

Corrigendum v1.0: Paper Two training protocol

STATUS: OWNER-RATIFIED FINAL - PRE-ANCHOR

Evidence checked: 2026-07-23

Affected anchored artifact

The Readout Indexes, the Scaffold Drives: Causal Asymmetry in Post-Grokking Dialects

- SHA-256: f1b992a8ff36abbce748bb0c0f113254884a0a43047a5c080923e8502e05839b
- Sepolia: 0xd2291414897586161edc5a7cc6ef20bf62aef54677baf12f2496c471be196dc6
- Sepolia block: 0xa77a46
- Block timestamp: 2026-06-02T16:43:48Z
- Anchor date: 2026-06-02, verified directly from the Sepolia block timestamp

The original PDF remains the anchored artifact. This corrigendum supplements it; it does not replace or silently revise those bytes.

1. Affected statements

The following statements in the anchored PDF are incorrect:

- Methods, page 11: the 217 models are described as reaching the grokking threshold within a universal 24,000-step training budget, and the population is said to span the same schedule conditions as Paper One.
- Appendix training grid, page 16: the table again gives a 24,000-step budget and presents the four extension cyclic configurations as the full-population grid.
- Page 17: "All training configurations and the per-seed/per-schedule grokking rates are reproduced from Dixon (2026, sec. 3)."

The universal budget and full-population grid are not reproduced from Paper One. Paper One §3.3 distinguishes schedule-dependent budgets and the original and extension cyclic grids.

2. Evidence basis

The replacement protocol was cross-checked against:

- anchored Paper One, SHA-256 6bf750296204d0a92bad77d02adb1b5682e56da36612861fe7384801f2d1f9b0, especially §3.3 (page 5);
- original generating script `exp9_full.py`, SHA-256 28d5a0f94a64a84e835c2cf61084097cb6c694238e765ee91165b599ffe93b9f. Lines 109-112 define the eight-run seed-42 noise-floor block at $\lambda = 1.0 + \epsilon$, $\eta = 0.01$, for ϵ in $[0.0, 1e-4, -1e-4, 2e-4, -2e-4, 3e-4, -3e-4, 5e-4]$;
- extension generating script `exp9_extra_seeds.py`, SHA-256 5722ed10d5516d7aca832977ae6c5db4eedb613552cfff94632fb5f6d13d33ec; and
- surviving extension bundle `exp9_extra_states.pt`, SHA-256 267349abbe2c1ab7eaa5b0354edca9c61070af731e74df1de50670a1c8654971, whose saved histories end at step 5,999.

No generating script or preserved history in the current evidence record uses a 24,000-step budget.

3. Correct population protocol

Batch	Seeds	Constant runs	Cyclic runs	Attempted	Training budget
Original	7, 13, 21, 42, 99	98	24	122	constant: 8,000; cyclic: 2,000 warmup + 3 periods
Extension	101, 137, 211, 313, 401	90	20	110	6,000 for both schedule families
Total	10 seeds	188	44	232	schedule- and batch-dependent

The original cyclic totals are:

- period 1,500: $2,000 + 3 \times 1,500 = 6,500$ steps;
- period 3,000: $2,000 + 3 \times 3,000 = 11,000$ steps; and
- period 6,000: $2,000 + 3 \times 6,000 = 20,000$ steps.

The schedule grids are:

- constant grid in both batches: λ in {0.3, 0.5, 1.0, 1.5, 2.0, 3.0} crossed with η in {0.005, 0.01, 0.02};
- original-only noise-floor block: eight additional seed-42 constant runs near $\lambda = 1.0$, $\eta = 0.01$;
- original cyclic grid: [0.3, 3.0] and [0.5, 2.5], each crossed with periods 1,500, 3,000, and 6,000, for seeds 7, 21, 42, and 99; seed 13 has no original cyclic runs; and
- extension cyclic grid: [0.3, 2.5] at periods 1,500, 3,000, and 6,000 plus [0.5, 2.0] at period 3,000, for all five extension seeds.

The resulting population arithmetic is 232 attempted runs, of which 217 met the test accuracy ≥ 0.95 grokking filter: 117 from the original batch and 100 from the extension batch.

4. Replacement language

Replace the affected Methods paragraph and Appendix table with:

The experimental population contains 217 grokked models selected from 232 attempted runs across two historically generated batches. The original batch used 8,000-step constant runs and cyclic runs consisting of 2,000 warmup steps plus three schedule periods (6,500, 11,000, or 20,000 total steps). The extension batch used 6,000 steps for both constant and cyclic runs. Both batches used the same constant weight-decay/learning-rate grid, but their cyclic grids differed as specified in this corrigendum. The original batch also included eight additional seed-42 noise-floor runs. Selection into the analyzed population used the same test accuracy ≥ 0.95 threshold within each run's applicable budget.

Replace the page 17 attribution with:

Paper One §3.3 documents the full schedule-dependent protocol for both the original and extension batches, including the extension batch's 6,000-step budget and distinct cyclic grid. The universal 24,000-step budget and single cyclic grid stated in this paper do not reproduce that source. The corrected batch-specific protocol is restated above.

5. Terminology boundary

Paper Two defines "function preservation" operationally on page 12 as held-out test accuracy after intervention. It does not establish equality of the real-valued logit function. Future summaries should therefore use "accuracy-preserving under the tested criterion," "decision-preserving on the held-out set," or "preserves grokking," except when referring to an exact logit-preserving reparameterization.

Scope

This is a methods-description and terminology correction. It does not revalidate every 217-model result or resolve the batch confounding in the raw PC1 fold split. The documented differences in batch, seed set, budget, cyclic grid, and logit scale remain collinear in the historical population.

Publication and integrity conditions

Upon publication:

- 1 this finalized corrigendum receives its own SHA-256, OpenTimestamps proof, and dated chain anchor;
- 2 the original f1b992a8... PDF remains available unchanged;
- 3 the paper's public card links both the original PDF and this corrigendum;
- 4 the public verifier checks the corrigendum's site file, OTS proof, and chain binding; and
- 5 no claim is made that a future corrected paper build is byte- or content-equivalent to the anchored original.

Review and ratification

- Version: 1.0
- Evidence checked: 2026-07-23
- Outside factual review: ACCEPT, with no required fixes, 2026-07-23
- Owner ratification: 2026-07-23

The outside reviewer assessed the complete drafts against evidence embedded in the self-contained review packet. The reviewer did not independently retrieve the anchored PDFs or generating scripts. The local pre-anchor verifier binds the source-artifact hashes used to assemble that packet.

The SHA-256 of the final rendered PDF is recorded externally in the pre-anchor manifest so the document does not attempt to contain its own hash.